U S G C R P

December 2000

United States Global Change Research Program

Subcommittee on Global Change Research

Committee on Environment and Natural Resources

HIGH-END CLIMATE SCIENCE: DEVELOPMENT OF MODELING AND RELATED COMPUTING CAPABILITIES

A report to the USGCRP
from the
ad hoc Working Group on Climate Modeling

USGCRP

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **DEC 2000**	2. REPORT TYPE	3. DATES COVERED **00-00-2000 to 00-00-2000**
4. TITLE AND SUBTITLE **High-End Climate Science: Development of Modeling and Related Computing Capabilities**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **U.S. Climate Change Science Program,1717 Pennsylvania Avenue NW,Washington,DC,20006**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**		
13. SUPPLEMENTARY NOTES		
14. ABSTRACT		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**51**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

The U.S. Global Change Research Program

The U.S. Global Change Research Program was initiated in 1989 as a high-priority effort to address key uncertainties about changes in the Earth system, both natural and human-induced; monitor, understand, and predict global change; and provide a sound scientific basis for national and international decisionmaking on global change issues. Congress codified the USGCRP in the Global Change Research Act of 1990, in order to provide for:

> ...development and coordination of a comprehensive and integrated United States research program which will assist the Nation and the world to understand, assess, predict, and respond to human-induced and natural processes of global change.
>
> ...increasing the overall effectiveness and productivity of Federal global change research efforts.

Since its inception, the USGCRP has been directed toward strengthening research on key scientific issues. The Program has supported research that has led to substantial increases in knowledge, improved predictive understanding, and documented evidence of global environmental change, including major scientific advances in the understanding of stratospheric ozone depletion, the El Niño-Southern Oscillation phenomenon, global climate change, tropical deforestation, and other issues.

The interagency Subcommittee on Global Change Research of the Committee on Environment and Natural Resources, a component of the National Science and Technology Council, provides overall direction and executive oversight of the USGCRP. Within this framework, agencies manage and coordinate Federally supported scientific research on global change.

Subcommittee on Global Change Research Participating Agencies and Executive Offices

Department of Agriculture
Department of Commerce
    National Oceanic and Atmospheric Administration
Department of Defense
Department of Energy
Department of Health and Human Services
    National Institutes of Health
Department of the Interior
Department of State
Environmental Protection Agency
National Aeronautics and Space Administration
National Science Foundation
Smithsonian Institution

Office of Management and Budget
Office of Science and Technology Policy

High-End Climate Science: Development of Modeling and Related Computing Capabilities

A report to the USGCRP from the ad hoc Working Group on Climate Modeling

Richard B. Rood, Chair
National Aeronautics and Space Administration (NASA)
Goddard Space Flight Center

Jeffrey L. Anderson
National Oceanographic and Atmospheric Administration (NOAA)
Geophysical Fluid Dynamics Laboratory

David C. Bader
Department of Energy (DOE)
Pacific Northwest National Laboratory

Maurice L. Blackmon
National Science Foundation (NSF)
National Center for Atmospheric Research (NCAR)

Timothy F. Hogan
Department of Defense (DOD)
Naval Research Laboratory

Patricia K. Esborg
Organizational Consultant/Strategic Change Management

U.S. Global Change Research Program
Subcommittee on Global Change Research
December 2000

FOREWORD

The continued development and refinement of computational models that can simulate the evolution of the Earth system is critical for the U.S. Global Change Research Program. Such models can be used to look backward to test the consistency of our knowledge of Earth system forcing and response over time periods of interest, and forward for calculating the response of the Earth system to projected future forcing. Models are exceedingly valuable tools for the development of scientific understanding. They also are critical to the environmental assessments used to synthesize Earth science results and provide information to decisionmakers.

Modeling has been an important part of the USGCRP during the last decade. Significant progress has been made in the development and application of atmospheric chemistry, ecosystem, and climate models. However, the U.S. system of distributed centers for modeling and supercomputing has produced both benefits and problems, especially for climate modeling. The system has been very supportive of creative approaches and the use of high-end modeling as a tool for discovery-driven scientific research. But it has not facilitated the development of the kind of product-driven modeling activities that are especially important for making climate model information more usable and applicable to the broader global change research community.

Over the past several years, a number of internal and external analyses have identified significant problems with the U.S. climate modeling effort. A 1998 report by the National Research Council, *Capacity of U.S. Climate Modeling to Support Climate Change Assessment Activities*, concluded that, while the U.S. community is "a world leader in intermediate and smaller climate modeling efforts, it has been less prominent in producing high-end climate modeling results, which have been featured in recent international assessments of the impact of climate change." In the modeling community, the United States has been falling behind several other nations in its ability to perform long-term climate simulations. The NRC report called for a nationally coordinated strategy and greatly improved computing capabilities.

As a step toward defining a U.S. climate modeling strategy, the Environment Division of the White House Office of Science and Technology Policy (OSTP) in January 2000 commissioned an *ad hoc* Working Group on Climate Modeling Implementation and charged it with preparing a plan for USGCRP climate modeling activities. The Working Group was chaired by Dr. Richard Rood of the National Aeronautics and Space Administration and included a representative from each of the Federal agencies with a significant investment in global weather and climate modeling. Their report and recommendations to the USGCRP are

contained in this document.

The USGCRP is in the process of drafting a new long-term Strategic Plan. Parts of the plan will address the issues of modeling capabilities and research on climate variability and change, and will put forth research goals and modeling objectives for the next decade. This report on the development of high-end climate modeling and related computing capabilities raises key issues for the program. We express our gratitude to the Working Group for their important contribution.

D. James Baker, Chair
Subcommittee on Global Change Research

CONTENTS

I. EXECUTIVE SUMMARY

1) Background

In January 2000, the Environment Division of the White House Office of Science and Technology Policy (OSTP) asked Dr. Richard Rood of the National Aeronautics and Space Administration (NASA) to form an *ad hoc* Working Group on Climate Modeling Implementation. The impetus for this request was the need to define and implement a strategy for climate modeling in the U.S. that responds to unmet National needs in climate prediction, climate-science research and climate-change assessment. The composition of the Working Group is given on the title page. There is a representative from each of the Agencies with a significant investment in global weather and climate modeling. The scientists in the Working Group have broad experience in climate and weather science, including atmospheric chemical modeling. In addition, the authors have experience in research and operational activities, high-end computing, and scientific and programmatic management. The Working Group also includes a sociologist who is expert in human systems and organizational change management. OSTP charged the Working Group to prepare a plan for climate modeling activities to serve as advice to OSTP and the US Global Change Research Program (USGCRP). The Working Group's report *High-End Climate Science: Development of Modeling and Related Computing Capabilities* is summarized below.

2) Summary of Findings

- The requirements and expectations placed on the climate community have grown to the point that the U.S. requires the service of a dedicated organization, which is referred to here as the Climate Service.
- The Climate Service must operate as a product-driven research organization. This is in contrast to the discovery-driven research that is predominant in U.S. science programs.
- A successful product-driven Climate Service requires leadership, management, and business practices that are substantially different from those used in discovery-driven research activities. The following attributes are required:
  - Clear definition of mission.
  - Executive management with the responsibility of overseeing quality control and delivering the climate products.
  - Unifying incentive structure that connects individual's activities with organizational goals.
  - Supporting business practices.
- There are three fundamental issues that provide complex and conflicting challenges to the formation of a Climate Service.
  - The high-performance computing industry has fundamentally changed. While this has provided better computational resources to many individual researchers, those applica-

tions that require the highest level of computing are struggling to remain viable. The tension is heightened by U.S. policy on supercomputers.

  - There are not enough people to provide either the scientific or information technology expertise needed to sustain all of the U.S. climate-science activities that strive to provide comprehensive capabilities. Key positions are going unfilled and students are not being trained to fill either the scientific research positions or the esoteric niches of software engineering, computational science, and computer science required for a successful high-end climate capability.
  - The multi-agency culture that developed to support the discovery-driven research activities is not well suited to support a more product-oriented climate service. A multitude of sub-critical activities reside in the different agencies, and there is no straightforward mechanism to allow concerted concentration of these resources towards common product-oriented goals.

- The Climate Service must be cognizant of and responsive to foreign centers that are defining the state-of-the-art in assessment and simulation capabilities and, increasingly, in scientific research.
- Issues related to high performance computing:
  - Shared-memory, vector computers manufactured in Japan, and essentially unavailable to U.S. researchers, have a combination of usability and performance that gives them far more capability than computers available to U.S. scientists.
  - Parallel computers manufactured in the U.S., often with distributed memory, are difficult to use. In addition, there are intrinsic limitations to the ability of climate-science algorithms to achieve high levels of performance on these computers.
  - Japanese-manufactured computers already delivered to foreign centers assure that U.S. scientists will have significantly less computational capability for at least three to five years.

    With the delivery of the next generation of Japanese computers, and continuation of current approaches to computing in the U.S., the gap between the U.S. and foreign centers will increase and exist for longer than five years.

    The purchase of Japanese vector computers would have an immediate impact on climate and weather science in the U.S. and offers the only short-term strategy for closing the computational gap between U.S. and foreign centers.
  - There is insufficient investment in the U.S. in software. A software infrastructure must be built to support both climate and weather activities. The software infrastructure must:

Facilitate the interactions of scientists at different institutions, allowing concurrent development in a controlled environment.

Facilitate the interactions of climate scientists and computational scientists, allowing more robust use of computational platforms.

Include development of systems software necessary for the operation of the hardware platform.

  - The U.S. policy requiring the use of distributed memory, commodity-based processor parallel computers increases the size of the needed software investment.

    Japanese vector computers require substantially less expenditure on software.

    The risk is high that software developed for U.S.-available computers will not achieve the performance and reliability realized by that using Japanese-manufactured vector computers.

  - Without the development of successful software, the deployment of large U.S.-manufactured hardware systems to increase computational capability is not justified.
  - The development of U.S. computational platforms for the Climate Service is a research activity and the research must be driven by the climate applications rather than by technological development. As a research activity, the intrinsic risks are high.

- Issues related to the shortage of human resources:
  - In order to focus adequate climate-science expertise for the Climate Service, a multi-agency response is needed.
  - Timely development of a Climate Service requires participation of presently existing capabilities.
  - Integration of efforts across institutions and disciplines is needed to achieve critical concentration of expertise on priority problems.
  - Competition for human resources with the mainstream information technology industry is high, and it will be impossible to populate the information technology staffs of multiple comprehensive climate-research centers.
- Issues related to existing multi-agency culture:
  - The current management and review process rewards individual accomplishments and tends to fragment efforts rather than focus them towards common goals.
  - The reward and incentive structure that currently exists is not strong enough to allow coordinated, product-oriented goals to rise to a level to be competitive with internal Agency missions and programs.
  - Fundamentally new management and business strategies are needed to support the product-driven Climate Service.

- The difficulties of facing these management issues are large and suggest that the initial implementation of the Climate Service should be as simple as possible.

  This is in conflict with need to integrate activities across institutions and disciplines to address human resource issues, to maintain similar levels of comprehensiveness as foreign centers, and to keep up with scientific evolution.

- The management issues require more directed authority and decision making than is possible within umbrella organizations, like the USGCRP, which were designed to guide research rather than generate products.
- Without addressing these management issues, providing additional funds to the existing programs will not be effective in the development of the Climate Service.

<u>3) Summary of recommendations</u>

- Formation of a Climate Service:
  - A Climate Service with a well-defined mission should be chartered to deliver simulation and related data products for understanding climate processes and predicting future states of the climate system.

    Built upon existing expertise.

    Clear separation of Climate Service functions from current Agency obligations.

    Not be located or assigned to any Agency or Center within the current multi-Agency framework.

  - We propose that an independent service, which is a concerted federation of the appropriate current agency capabilities, should be formed. The existing agencies need to act like member states, drawing from a concept successfully used in the European Union.
- Management and Business Practices:

  Without a new business model incremental funding of existing organizations will not provide needed capabilities. The Climate service requires:

  - An integrating management structure.

    Executive decision-making process.

    Supporting incentive structure.

  - Supporting business practices.
  - Appropriate types of external review and oversight process.

    Stability

    Insulation from short-term programmatic volatility

- Computational Resources:

  The Climate Service requires:

  - o Dedicated computational resources with the highest level of capability.

    Computational resources must be:

    aligned with the generation of the Climate Service products (*i.e.* application driven).

    under the management of the Climate Service.

  - o U.S. policy on high performance computing adversely affects the Earth sciences.

    This increases both the expense and risk associated with climate science.

- Number of centers / integration:

  - o We recommend two major core simulation activities.

    The first is focused on weather and should build from the National Weather Service.

    The second is focused on climate, and its definition requires successfully addressing a number of the strategic and organizational issues discussed throughout this document.

    - The decisions on what should be included in a nascent climate service, e.g. seasonal-to-interannual, greenhouse scenarios, chemistry, data assimilation, etc., are among the most difficult to reconcile. There is a need to integrate activities across institutions and disciplines to address human resource issues, to maintain similar levels of comprehensiveness as foreign centers, and to keep up with scientific evolution. This is in conflict with the difficult management challenges that suggest the initial implementation of the Climate Service should be as simple as possible. The complexities of the integration issues are beyond the scope of the current deliberations.

    - It is critical that initial steps be made to develop a credible and competitive high-end climate capability, and we are concerned that potential agency and political positioning over the location and running of a potential Climate Service will delay its formation.

    The Weather Service and the Climate Service should undertake the development of the formation a unifying infrastructure to allow effective transfer of expertise and algorithms.

- Size and budget of core simulation capability for Climate Service

  - o On the order of 150 scientists, software engineers, and application-directed computational scientists, programmers, and computer scientists need to be dedicated to the modeling capabilities of the Climate Service.

- The total funding for the modeling and computing component of the Climate Service is on the order of $50 M.

  There are large uncertainties in this number because of computational policy issues that are beyond the scope of the climate-science community. The $50 M is a lower limit.

4) Final Comments

The details of implementation of the Climate Service will require significant planning and be dependent on a number of interrelated decisions that must be made by the Agencies. Strong leadership is required, both within the Agencies and at a level higher than the Agencies. The implementation can and should be incremental. In fact, we believe that with the definition of a stable vision there are a number of existing activities that could form the core of a future Climate Service. There are already moves by all of the Agencies to better integrate and unify modeling and computational activities. If these can be orchestrated towards a long-term vision, then substantial steps can be taken while the details of the Climate Service are developed and evolved. Again, without a new business model and management strategies within which to organize the Climate Service, there is a danger of simply rearranging the current activities, which will not be successful.

Finally, we emphasize that is artificial to speak of a climate-science capability, a national climate service, without integration of modeling and data (*i.e.* observational) activities. As charged, we addressed the data activities, but they were not explored in as much depth as the modeling activities. We state, explicitly, that many of the same underlying problems affect the environmental data undertakings of the U.S. as affect the modeling community, and integrated, systematic solutions are ultimately needed. Additional funding is crucial to both develop foundation climate observing systems and to integrate and maintain existing data sets for climate applications.

II. CHARGE: Climate Modeling Implementation Plan Working Group

The Working Group should develop a draft implementation plan for US Global Change Research Program climate modeling activities. The draft plan should be presented to the Subcommittee on Global Change Research by May 1, 1999. The plan should cover all USGCRP climate-modeling activities, and should address two time frames: FY2001-2005 and FY2006-2010.

Two options should be developed:

Option 1: Assume level funding of USGCRP climate modeling activities (at about $70M per year). Full funding of Administration's IT initiative leads to development of more capable supercomputing facilities and availability of competitively awarded software development funding. Enhanced cooperation between climate modeling and IT.

Option 2: Assumes possibility of incremental funding increase of $10M per year for USGCRP climate modeling activities over 5 years (from about $70M per year to about $120M per year). Full funding of Administration's IT initiative leads to development of more capable supercomputing facilities and availability of competitively awarded software development funding. Enhanced cooperation between climate modeling and IT.

The working group should, at a minimum, address the following set of questions.

Issue Identification.

What are the climate and weather modeling requirements of the U.S.? What improvements are needed in current capabilities and why?

What is the correct balance between software and hardware funding to achieve current requirements and necessary improvements?

What are the relevant weather/climate modeling activities now being pursued by the research and operational community (e.g., what spatial scales and types of models should be included)?

What are the science, IT, and human resources difficulties now hindering these efforts and what are the possible science/IT/human resources solutions to these problems?

What are the emerging science and IT challenges? For instance, how should modeling change in response to the very large increase in the amount and quality of observations data expected in the next ten years?

What would be a realistic timeline to achieve these solutions?

Management Approach.

Should US weather and climate modeling efforts be more closely integrated?

If so, what management structure should be developed to ensure that these efforts (weather, climate) are effectively integrated? Are modifications to the current "distributed" US approach warranted?

How can weather and climate modeling efforts best be integrated into the broader IT effort?
What types of IT resources are likely to be available in the next ten years, and how does climate modeling need to change to best take advantage of such resources?
How can we ensure that the most cost-effective IT innovations are adopted and that the IT resources are properly allocated across the related science/IT communities?
How can we ensure that common modeling and IT tools are developed and used effectively across the key modeling groups?
What steps are needed to stimulate more effective interaction between climate modelers, researchers, and the larger IT community?
How do we deal with the human resources issues (how can we get more stable funding for modeling groups)? Is it possible to engage smart young IT people in climate modeling over the long term?
What steps are needed to ensure that climate modeling is prepared for the massively parallel architectures that are becoming predominant in the computing world?

High-End Climate Science:
Development of Modeling and Related Computing Capabilities

1) Purpose

The United States requires the capability to deliver state-of-the-art products for weather forecasting, seasonal-to-interannual predictions, and climate assessments in order to answer policy-relevant questions in a timely and efficient way. This includes the development of model and data systems that provide the best possible direct information on how environmental factors impact the U.S. and its assets, as well as evaluation of the relationship of the country's regional environment to global environmental processes.

This report will outline an approach for the U.S. to develop a high-end climate modeling capability to address the strategic interests of the Nation. These interests range from the benefits that would be received by improved model predictions of weather and seasonal characteristics to the increasing use of environmental assessments in the determination of regional and global policies. It is critical that U.S. policy makers have at their access the ability to best evaluate, for example, the impact that changes in the emission of a particular greenhouse gas might have on U.S. interests. Such a calculation must be timely to a particular deliberation and would require an ensemble of comprehensive scenarios to be run and evaluated in a matter of months. Presently, this calculation would require a number of years. Since several other countries have the capability to make these assessments at a rate many times faster than the U.S., the U.S. is at a strategic disadvantage. A number of shortcomings in the U.S. approach to achieving a high-end climate science capability are documented in the report *Capacity of U.S. Climate Modeling to Support Climate Change Assessment Activities* (National Academy Press, 1998). These shortcomings are tangibly indicated by the U.S. National Assessment, which relies on results of Canadian and British models because no U.S. model met the criteria established for inclusion in the assessment. (1)

Our report first provides the context of the current state of climate-science in the U.S. and then identifies a number of key issues that must be faced if the U.S. is to develop a climate service with a credible high-end modeling capability. We assert that a capability that is state-of-the-art, comparable to the best in the world, is in the best interest of the Nation. Therefore, evaluation of U.S. capabilities relative to those of other countries is an important metric of success. We then subdivide climate-science research activities, distinguishing between discovery-driven and product-driven research and defining high-end modeling. After stating the need to build a product-driven climate service, we explore the issues that must be addressed to build this service by first splitting climate-science into three Elements: Modeling, Data, and Computational Systems. These Elements should not, however, be treated separately, and we discuss the need for integration across these Elements as well as across traditionally separated disciplines such as weather forecasting, seasonal prediction, data assimilation, atmospheric chemistry, and coupled climate modeling. Following that, the software and hardware aspects of computational systems are discussed. Then issues of human resources and the characteristics of an institution to deliver the required climate-science products are presented and concluded to be as challenging and important as any scientific and technical issues that confront the field. Finally, general guidelines for investment and management strategies are made.

2) **Current situation**

The benefits of weather forecasting to both civil and military enterprises of the United States are well established. As the capabilities of the weather and climate modeling community have increased, the expectations for products have expanded. Now, there are societal needs for seasonal and interannual predictions, environmental assessments, and sophisticated model-assimilated data sets to aid in the quantification of the physical and chemical processes that determine the balance of key environmental parameters (*i.e.* temperature, water, winds, ozone, etc.). Historically, many of the capabilities to meet these needs have grown out of successful research activities that have expanded to take on the responsibility to provide these products.

A number of factors external to the research community have changed in recent years that stress these capabilities. In fact, the report *Capacity of U.S. Climate Modeling to Support Climate Change Assessment Activities* (National Academy Press, 1998) concludes that there are serious deficiencies in the country's ability to provide the necessary products for climate assessment. While the general quality of climate science in the U.S. remains competitive with any in the world, there is little doubt that the ability of centers in Europe, Canada, Japan, and other countries to perform computational experimentation far exceeds that of their U.S. counterparts. Foreign centers are able to run more comprehensive models, use more observations, and increasingly capture the intellectual interest of the extended scientific community. In fact, observational capabilities developed by U.S. agencies and breakthroughs by scientists at U.S. research centers and universities are often being more effectively exploited abroad.

There are a series of complex problems that must be addressed if the capabilities of U.S. climate science are going to be boosted to meet the expectations of the Nation. The solutions to these problems are not straightforward and are entangled with a number of issues that are outside of the direct control of climate scientists or the managers within the agencies that fund climate science. There are three major problems that sit at the foundation of the crisis currently facing the U.S. climate-science enterprises.

I. The high performance computing industry has fundamentally changed. While this has provided better computational resources to many individual researchers, those applications that require the highest level of computing are struggling to remain viable. The tension is heightened by U.S. policy on supercomputers, which leaves U.S. scientists with a significant usability and performance gap compared with their foreign counterparts.

II. There are not enough people to provide either the scientific or information technology expertise needed to sustain all of the U.S. modeling and assimilation projects that strive to provide comprehensive capabilities. Key positions are going unfilled and students are not being trained to fill either the scientific research positions or the esoteric niches of computational science and software engineering required for a successful high-end climate capability.

III. The multi-agency culture that developed to support discovery-driven research activities is not well suited to support a more product-oriented climate service. A multitude of sub-critical

activities reside in the different agencies, and there is not an effective business plan with a consistent incentive structure to allow concerted concentration of these resources towards common product-oriented goals.

In addition to these three major problems, the growth of high-quality climate-science capabilities outside the U.S. places the U.S. climate-science community in a difficult position. Line managers at all U.S. institutions feel the pressure to compete, especially with the European Center for Medium-range Weather Forecasts (ECMWF), The Hadley Center, and The Max Planck Institute for Meteorology (Hamburg). In many cases weather and climate products from these centers are held up as the standard, and there is little hope of the U.S. meeting these standards in the immediate future. Thus, the public and both government and private sector institutions are deprived of the benefits of the best possible capabilities. Further, as model assessments of environmental changes and their impact play a more important role in determination of a wide variety of policy decisions, simulations provided by non-U.S. centers lay the foundation on which these decisions are made. It is in the self-interest of the U.S. to be able to generate state-of-the-art environmental modeling and model-assimilated data products.

These three major problems, coupled with the competitive pressures that come from many non-U.S. centers, place the climate-science community in an unprecedented situation. Simply adding funds to the existing organizations will not yield a cost-effective, robust, and flexible high-end climate-science capability. The problems are not purely technical, but require addressing sociological and business problems that sit at the foundation of the U.S. multi-agency scientific culture. Solutions are not straightforward and have significant risk of failure. Computational issues are at the center of this risk, but the sociological and business issues are as daunting and important. This report will expose the elements of these issues more fully and then suggest strategies with which to confront these issues.

3) Scope of Document/Underlying Definitions and Assumptions

The charge for this document (attached to Executive Summary) focuses primarily on climate modeling and those activities linked to the U.S. Global Change Research Program (USGCRP). Questions are raised about links to programs that focus on weather forecasting and seasonal-to-interannual predictions. Further questions are related to the “data” missions of the U.S. agencies. The report will focus on the climate-modeling problem with the following constraints.

1) It is artificial to speak of a climate-science capability, a national climate service, without integration of modeling and data (*i.e.* observational) activities. As charged, we will discuss the data activities, but they will not be explored in as much depth as the modeling activities. We state, explicitly, that many of the same underlying problems affect the environmental data undertakings of the U.S. as affect the modeling community, and integrated, systematic solutions are ultimately needed.

2) Climate modeling and weather modeling programs, and their related data programs, have developed more or less independently in the U.S. There is significant overlap between the two

applications and more thorough integration is needed. The fact that in Europe operational weather models stand at the foundation of their climate modeling activities is frequently viewed as a strength. There are, however, a number of differences between the U.S. and Europe. Since the U.S. has historically invested many dollars in both climate and weather forecasting, a number of options exist – an integrated weather and climate capability could be developed from either a traditional weather or a traditional climate foundation. Reasonable arguments can be made for both of these options. We propose that on a five to ten year horizon, it is in the best interest of the U.S. to look towards the integration of weather and climate activities, with the maintenance of one or two integrated core efforts to provide the needed weather and climate services.

3) In the U.S. a number of groups have been formed to orchestrate and guide research programs towards strategic goals. The USGCRP and the U.S. Weather Research Program (USWRP) are such groups. As these groups address the problems of their individual communities, they advocate initiatives that potentially come into conflict with each other. For instance, they compete for the same human resources within a stressed resource base. If these activities are, themselves, not integrated, then they will contribute to the fragmentation of the country's capabilities.

3.1) Changes in Research Needs and Expectations

Historically, scientific inquiry into the Earth's climate has been primarily maintained by discovery-driven research. The structure for funding of U.S. science was developed to support this research and focuses on the activities of individuals or small groups that can be managed with relatively little formal management structure. The net result of this history is effective, high-quality scientific inquiry.

Presently, climate scientists are seeing a significant change in the demands made on their scientifically derived results. New expectations range from seasonal predictions to provision of best, state-of-the-art at a given time, assessment of potential changes in the Earth's climate, including related issues of chemical composition and regional pollution. Traditionally, these extended activities have been addressed by taxing discovery-driven research activities. This approach was adequate when the needed capabilities of these extended applications were exploratory, and when U.S. scientific activities were at the leading edge of international science activities. This is no longer the case.

The current situation of climate science in the U.S. is one of great tension between expectations and the capability to meet those expectations. The strength of U.S. efforts is diversity, with many individual scientists having unprecedented resources. They control their own computational resources and have desktop or departmental capabilities that allow flexible, thorough experimentation. The line managers at Agency laboratories, however, function with great frustration. They do not have the infrastructure to build the algorithms and then deliver the products that are expected of them. They do not possess all of the needed expertise in their organizations. They do not have the ability to incorporate, effectively, the diverse expertise that resides in the community. They do not have the resources or authority to integrate the diverse activities towards focused goals. And, finally, if they

seek the needed changes to remedy the situation, they meet with great resistance because the needed changes challenge the culture of U.S. science.

3.2) Definition of Discovery-driven Research

Discovery-driven (2) research is closely associated with individual Principal Investigators (PI's), who propose to investigate particular hypotheses or phenomena. In terms of deliverables, discovery-driven science strives to deliver "understanding" with the tangible product being research reports, related scientific presentations, and ultimately, peer-reviewed journal manuscripts. The customer is both the peer community in which the PI operates and the Program Manager who funds the research. In most cases, the delivery of peer-reviewed publications leads to a statement of customer satisfaction or dissatisfaction realized through peer-reviewed proposals that help to determine whether or not a particular effort continues to receive funding. Time-criticality of the delivery of discovery-driven research is determined by the evolution of the field and the requirement that the PI remain current and competitive in the community of peers. Extension to larger group activities is through *ad hoc* collaborations that develop through either mutual self-benefit or steering by program or center management.

3.3) Definition of Product-driven Research

In product-driven science, the primary deliverables are data products, simulations, and algorithms that are used by others in a variety of applications. In many instances, the applications contribute to discovery-driven research leading to the generation of peer-reviewed papers, which then testify to the quality of the original product. Increasingly, however, the products are used to support societal and policy needs. Direct customer feedback requires quickly addressing product shortcomings, which precludes the more contemplative time scales of discovery-driven science. Ideally, peer-reviewed papers stand at the basis of the scientific integrity of the product, but technical reports and documentation become more important. Product performance and customer perceptions of the scientific quality of the product can overwhelm the role of peer-reviewed papers. Scientific inquiry is often cut short in order to provide the closure needed to assure product delivery; the customer effectively takes away some of the Principal Investigator's prerogatives. All told, product-driven science requires resource investment in the infrastructure to support the product, and directs scientists away from peer-reviewed papers delivered on a more or less casual schedule.

3.4) Definition of High-end Modeling

This document proposes an integrated approach to climate science. By charge, the emphasis is on high-end modeling and its related computing. We assert that a balanced, strategic approach is needed, where the word "strategic" will be used to suggest activities that require any or all of the following

- combination of the Nation's resources
- development of specific capabilities needed to meet the Nation's goals

- investment of resources prior to a specific requirement for a particular activity
- consideration of time scales that are longer than the life of particular projects and tasks

The word "balanced" is used to make explicit that resources must be invested in all of the elements necessary to develop the required climate-science capabilities. This requires a systems approach where the capabilities and interactions of all the individual elements are considered with the goals of the combined capabilities taking priority over the goals of the individual elements.

While the emphasis of this document will be on modeling and its related computing, there is full recognition that issues of data collection, data use, and data management are a key part to climate science and to a successful modeling capability. Many of the concepts introduced here are relevant to the Data Element (Section 4) of climate science, but they will not be developed to the same level as the issues regarding modeling.

The following pairs of adjectives can be used to divide research activities broadly (and imperfectly) into two categories.

Column I	Column II
product-driven	discovery-driven
multiple investigators, teams	single investigators, small group
multiple processes	single process
self-determining	mechanistic
large	small

Modeling capabilities described by the adjectives in Column I are the primary focus of this document and will be defined as "high-end modeling." (3) Current high-end modeling activities include those used to forecast weather, to provide seasonal predictions, and to provide both chemical and climatic assessments on longer time scales. High-end modeling activities also include related data assimilation activities, which are central to chemistry and climate studies. This document focuses, by charge, most strongly on activities associated with the U.S. Global Change Research Program (USGCRP); however, an essential aspect of a comprehensive strategy includes increased formalism of the interactions between climate-oriented activities (longer time scales) and more operationally oriented forecast activities (shorter time scales).

Efforts in both Column I and Column II are both essential parts of a vital modeling capability. The current balance in the U.S. is tipped towards Column II and this document will identify the necessary steps that must be executed in order to grow an effective high-end modeling capability. This capability cannot be built by redirection of research funds that are linked with the USGCRP and will require infusion of new funds. The "mission" agencies, NOAA, NASA, and DOE, already fund some activities that are product oriented. NSF, through the Community Climate System Model Program, has taken on the responsibility to provide a research facility for the climate community. These existing capabilities could contribute to the core of a putative climate service. This requires the Agencies to clearly distinguish the funds that might support this core from those funds that support their discovery-related research and manage them appropriately. With the consideration of these

existing product-oriented activities, which are broader than those just associated with USGCRP, the needed capabilities can be built from a combination of existing and new funds. The new funds must support the development of product-driven activities and not simply appear as enhancements to existing research activities. Finally, the current discovery-driven research programs generally support high quality and important investigations; therefore, to build a product-driven research institution at the expense of the discovery-driven programs would undermine the underlying research environment that is broadly cited as the greatest strength of the U.S.

3.5) Formation of a Climate Service

The modeling, data, and computing capabilities need to be brought together in a Climate Service, an organization charged with the mission of providing necessary climate products. The formation of this Climate Service requires the development of aligned business practices to support the integration of currently dispersed resources. It is a difficult problem, and stands in conflict with much of the current culture that supports discovery-driven research activities. There are existing activities that, with little controversy, might provide some of the fundamental components of an evolving Climate Service. We assert that, in the near term, the vision of a Climate Service needs to be advanced and that these components be assembled, under strong stable leadership, in concert with this vision. We further assert that a metric of success of the Climate Service will be that it provides a resource to the discovery-driven research community. That is, the Climate Service will draw from the discovery-driven research community and the discovery-driven research community will not only benefit from the products of the Climate Service, but will also benefit from the intellectual interactions with the Climate Service.

4) Elements of Climate Science

In order to develop a strategic approach to build a national capability of high-end modeling, climate science will be broken into three large elements, Modeling, Data, and Computational Systems. It is safe to say that all of these elements currently command substantial monetary resources and are in varying states of health. In terms of total cost, the Data Element is the largest of the three. This reflects the cost of the development, deployment, and operations of the instrumentation needed to collect the observations, as well as the cost of the data and information systems that provide data services. The charge for this document focuses on modeling and its related computing, and the observational system issues are limited to the model-data interface. Again, a comprehensive national climate service requires thorough integration of the modeling and data activities; therefore, the charge of this document may have an artificial boundary that ultimately needs to be breached.

Relative guidelines can be given for the budget of the Modeling and Computational Systems Elements. A single institution with the charge to deliver high-end modeling products to a spectrum of customers should expect its expenditure in the Computational Systems Element (see Section 4.3) to be equal to or greater than its investment in scientific personnel to support the Modeling Element. Depending on how the single institution organizes itself, the ratio of cost is on the order of 1:1 to 2:1

(Computational Systems/Modeling). There are tremendous uncertainties in the estimation of budgets for computational systems because of the policies that govern high performance computing in the U.S.

Further, there must be recognition of the need to build interfaces of the high-end modeling institute with external communities. These external communities include: the customer community, the users of the products of the high-end modeling institute; the diverse community of discovery-driven researchers, whose activities must contribute to and benefit from the high-end modeling enterprise; and other modeling centers.

4.1) Modeling

The Modeling element refers to those activities involved in building, applying, and validating geophysical models. Historically, research activities in several agencies focused on models that represent the atmosphere, the oceans, the land-surface, and the cryosphere. Within these disciplines, models specific to particular problems were developed, for instance, numerical weather prediction models, stratospheric chemistry models, and long-term climate assessment models. Currently, there is a movement to view these discipline models as components of combined models that represent all of the importance processes of the Earth system. Therefore, leading-edge activities today are focused on coupling atmospheric, oceanic, land-surface, cryospheric, and, even, thermospheric models. In addition to these geophysically based discipline divisions, there are application-based divisions in which computationally demanding comprehensive models have been built. For example, for the atmosphere there are global-scale, regional-scale, mesoscale, and cloud-resolving models. The primary reason these divisions were made originally was to develop tractable problems that allowed quantitative paths to be developed. This incremental approach allowed scientists to organize the complexity of environmental phenomena; focus on specific important processes; and adaptation of problems to available computational devices.

The broad and substantial U.S. investment in this quilt of modeling activities developed both scientific expertise and model algorithms that were and are pioneering. In fact, as other countries have developed modeling capabilities, they have been able to directly benefit from these pioneering efforts. In effect, the U.S. funded much of the initial discovery-driven research that allowed, globally, the more efficient development of comprehensive capabilities by other countries. The U.S. is now faced with the need to develop a similar capability, requiring a new sort of infrastructure to facilitate this integrated activity.

We believe that models built and maintained by U.S. scientists are in some cases competitive with those run anywhere. However, the U.S. capability is fragile, and the evolution of U.S. capabilities is lagging substantially. Increasingly, U.S. competitiveness in environmental modeling rests on the success of past investments and the well-respected role of U.S. scientists in the international community. A major part of this fragility is directly related to computational resources, where the number of numerical experiments run with a particular model is much smaller than that being run with similar models overseas. This reduces the validation and analysis of model results, compromising the scientific process on which the existing U.S. strengths lie.

While the historical approach to funding modeling activities did provide a wealth of scientific results and algorithms, simply increasing the funding within this approach is not a viable path forward. There is too much fragmentation in the current research activities and the discipline- and agency-linked approach to modeling supports, and often increases, this fragmentation. The focus of effort must be raised out of the disciplines and turned towards the products that are expected from the models. It is the definition of the products and the management of the resources that allows decisions to be made about which development paths to follow.

Only with a systematic product-dependent approach can questions about the merits of investment in more resolution or more sophisticated physical parameterizations be weighed. Scientifically, increased resolution and improved physics both demand attention. However, the two are intimately related, with, for instance, increased resolution moving the development of physical parameterizations closer to first principles. Higher resolution with inadequate model physics will not lead to more credible results. Better physics in low-resolution models will still leave important questions unanswered, such as regional climate impacts. Only a comprehensive, systematic approach will suffice. Further, model development should also be linked to the Data Element, exploiting information in existing observations and directing the observational strategies to the most important new observations. Therefore a balanced approach is needed, and it is the development and nurturing of the expected products that provides the primary mechanism for guiding the balance. It is the job of the responsible managers to determine the balance while maintaining the integrity of the underlying science. As in any business, the responsible manager must consider the capabilities of other organizations, as their capabilities contribute to the definition of what the community feels is both credible and state-of-the-art.

4.2) Data

Climate observing systems have been addressed in a separate report, *Adequacy of Climate Observing Systems* (National Academy Press, 1999). That report follows from earlier findings that the Earth's climate observing system is "inadequate" and "deteriorating," and endorses a set of principles that need to be addressed in order to assure an adequate observing system. Parallel to the findings of the report, *Capacity of U.S. Climate Modeling to Support Climate Change Assessment Activities* (National Academy Press, 1998), there is the finding that the U.S. agencies do not have an integrated, systematic approach to provide the necessary climate observations. Therefore, many of the management issues that are discussed below are relevant to the Data Element of climate science. Indeed, the Modeling and Data Elements must be addressed in a systematic and integrated way.

The inadequacies of the data system come in the face of a number of apparent contradictions. In terms of quantity of data and parameters measured, there is more data than ever before. In addition, as more and more space-based instruments are launched the amount of data stands to overwhelm the capabilities of data systems, modeling systems, and the community to utilize the observations in a timely manner. As in the case of models (discussed in Section 4.1), this contradiction is embedded in an imbalance between discovery-driven and product-driven research. However,

the details of the Data Element are different.

The largest amount of money under the USGCRP umbrella is targeted towards discovery-based research measurements from space. Many of these observations target identification or quantification of specific physical and chemical processes. The data used to establish baselines of key parameters in climate studies come, however, from the operational data collected primarily in support of weather forecasting. Again, the greatest bulk of the operational data come from satellites. In total, it is safe to say that, worldwide, these data are underutilized, and there have been a number of workshops focused specifically on this problem (*e.g.* Errico, 1999). These operational data have not been collected with accuracy and stability requirements suitable for climate research.

While the bulk of the operational data are from space-based platforms, the most critical climate data sets are, still, from traditional land- and balloon-based observational systems. Many of these observing systems are simple technology. The collection of high-quality observations from these systems requires routine, careful process, with significant human oversight over both process and data evaluation. There is ever-present pressure to replace these observing systems with automated systems, or to force reliance on satellite instruments. Without careful planning, such changes threaten the fidelity of the climate observing system. In addition, given that oceanic and land-surface parameters are key climatic variables, there is the need to develop suitably based observing systems to provide the necessary foundation. Some focused capabilities have already been implemented, such as the Network for Detection of Stratospheric Change (4). Concepts for others have been developed, but require funding to build, deploy, and maintain them.

Ultimately, a critical subset of climate observations falls into the cracks between the requirements for operational weather prediction and the, often, high-technology observing systems designed for exploratory research. In the end, no Agency, no entity, has the responsibility with aligned resources, to collect and provide stewardship of climate observations. The well-intentioned efforts of the Agency research programs to address the needs for climate observations fall short.

4.3) Computational Systems

The computational requirements for climate-science modeling have traditionally been one of a handful of problems that easily and necessarily consume the capabilities of the most capable of computers. Indeed, progress in Earth science is dependent on computational resources. Computers are used to provide simulations, which are used not only for prediction but also for experimentation to isolate and understand underlying physical and chemical processes. Computers lie at the core of data assimilation, a process where observations are melded with model simulations to provide accurate estimation of the state of the atmosphere, oceans, and land-surface. Ultimately, computers provide the tool that allows the integration of the complexity of physical, spatial, and temporal scales into comprehensive models that represent our best expression of the behavior of the climate system. While we often think of “the computer” as the key component of computational systems, software is what is actually built by the scientific community. The software connects the intellectual endeavors of

the scientists with the enabling technology of the computer.

The software that comprises a comprehensive model of the atmosphere-ocean-land-surface system consists of many hundreds of thousands of lines of code. The analysis software in related data assimilation systems is of the same size. Many hundreds of functions and routines are involved. The routines change with scientific advancement, and there is a continual demand to improve the accuracy of, or eliminate, the many approximations that reside in all modeling and assimilation systems. There are not standard or unique ways to express, even, the relatively stable, non-controversial, components of these models; therefore, the human diversity of the discovery-driven research community is reflected in the software systems that have been developed over the past twenty years or more.

From the 1980s through the mid-1990s, the interface with the computational platform was relatively stable. However, with the move in the U.S. away from shared-memory vector computers to distributed-memory commodity-based processors (5), the computational environment has become volatile. The components that comprise the high-performance hardware available to U.S. scientists are now dependent on the rapidly changing commercial market. How these components are collected together to provide high-performance computing platforms is often a research activity unto itself. Since the market for high-end computing is both small and specialized, corporate interest to develop platforms to support high-end applications does not draw top priority. The net result of these changes is that much attention needs to be placed on the interface of the application software with the computing environment in order to provide successful strategies for code runability and performance.

From a computational perspective, software is of the highest priority. A hardware-centric approach is no longer useful and the notion that "one computer fits all" is naïve. Without a large and successful investment in software, purchasing U.S.-available, distributed-memory parallel machines will do little to advance the overall state of climate-science in the U.S. Since the software challenges are less daunting for the tightly integrated vector supercomputers provided by Japanese vendors (*i.e.* NEC, Fujitsu), their purchase would benefit the U.S. community immediately. In either case, however, an approach focused on sound principles of systems design and systems engineering is required. The planning for systems must extend beyond a focus on just the core models or assimilation algorithms to include adequate interfaces with data, aspects of semi-automated validation (quality assurance), and interfaces with customers. The successful development of a high-end climate-science-modeling capability will require the development of computing systems focused on this problem. The continued pursuit of technology-driven development of hardware capabilities, with the idea that climate-science applications will be able to benefit from these developments, provides, at best, an uncertain progression in the development of the needed climate-science capabilities.

Software and hardware systems must be developed with a definitive focus on the specific applications to be executed. The focus must move to the development of software systems that represent a state-of-the-art expression of the science and a flexible robust interface to an uncertain computational environment. The potential capabilities of supercomputing hardware are only realized with the development of successful supercomputing software. It is safe to conclude that the current dollar investment in software is inadequate, and that the challenges in software development are

enormous with many intrinsic risks. The issues of software design and development will be discussed more fully in the next section.

4.4) Integration across Elements, Institutions, and Disciplines

The traditional evolution of the U.S.'s diverse modeling activities was highlighted in Section 4.1. As the individual disciplines move to incorporate algorithms from other disciplines, program managers and funding officials expect the research in these different areas to contribute to research in other areas. However, the contribution of one research area to another is far from optimal, and it is one of the challenges of the climate-science community to make these connectivities more effective.

The need for more effective integration goes beyond programmatic efficacy. The quality of the different sub-disciplines has developed to the point that many scientific challenges now lie in expansion out of the original discipline. This is quite clear in weather prediction and climate simulation where many of the physical processes are essentially the same, and where reconciliation of long-term (*i.e.* climate) behavior with short-term (*i.e.* weather) behavior confronts fundamental processes. Similarly, reconciliation of the behavior of parameterizations across spatial scales from cloud-resolving to global obviously improves the robustness of predictive models (7).

The development of multiple, comprehensive, high-end capabilities, for example a unified chemistry-climate model, from the existing specialized efforts spreads resources thin. For instance, the extension of a stratospheric chemistry model to include tropospheric chemistry and then further to climate simulation requires the inclusion of physical processes that are already central to the activities of the traditional climate community. Similarly, the extension of a tropospheric climate model to study stratospheric processes and to further include ozone photochemistry requires the inclusion of physical and chemical processes that are core to the stratospheric chemistry community. There is a natural tendency for these groups to look at their native model as the core tool and then either adapt algorithms from the other community or invent derivative algorithms that adapt to the particular model. This leads to models that are not of uniform quality across their entire suite of algorithms. Further, it consumes the time and energy of scientists in reworking what are often routine algorithms and while it advances the completeness of the models, it often does not advance the scientific integrity of the field as a whole. The development of a unified infrastructure in which scientists from multiple institutions can contribute concurrently is an essential, and currently absent, ingredient in an effective high-end modeling capability.

Therefore, substantial integration is needed across disciplines, across institutions, and across the three Elements: Modeling, Data, and Computational Systems. In fact, the historical programmatic investment in each of these parts individually lies at the fragmentation that currently permeates the field. Recently, there have been some successes as scientists and program managers recognize the fragmenting processes and try to overcome them. However, the sociological inertia in the current culture is difficult to overcome. Suitable levels of integration need to be pursued as strategic initiatives, but they must be kept well enough defined to better the possibility of success (8).

5) Issues of Computational Systems

Adequate computational resources are at the core of developing a high-end modeling capability to meet the Nation's needs. Ten years ago, supercomputing was purchasable commodity, with the particulars of climate-science computation requiring only incremental adaptation, compared with other disciplines, to the computing environment. The U.S. policy on high performance computing moved to focus on distributed-memory architectures using processor elements that have a market base broader than scientific applications (9). In the U.S., the development of shared-memory computers with specialized vector processors, which had been the standard used for climate science for a number of years, ceased. However, the sustained development of the shared-memory vector technology continued in Japan (10). A major source of tension in the climate-science enterprises of the U.S. arises from the wide-spread access to the Japanese supercomputers by scientists in virtually every country but the U.S.

Many millions of dollars have been spent on the development of applications software to use the distributed memory computers available to U.S. scientists. The success of these developments has been mixed. Certain subsystems of an end-to-end application can be made to run very fast. However, complete end-to-end systems that consider the climate-science problems of data ingest, simulation, assimilation, quality assurance, diagnostics, and push of output products to customers are generally not successful (11). The uncertain success of U.S. endeavors is sharpened when the competitive aspects of climate-science activities are considered; namely, climate-science centers in many other countries have more complete models, use more observations in their studies, and produce more simulations and assimilated data sets than U.S. counterparts. Further, there is the perception that the best ideas from the diverse U.S. research activities are implemented more readily at non-U.S. centers. Bottom line: the usability of Japanese supercomputers is much higher than that of U.S. computers, and they are therefore, pragmatically, "faster."

The U.S. climate-science community is faced with a difficult, perhaps intractable, problem. With the present national strategic focus on distributed-memory computers, a tremendous expenditure on software is required. This includes not only applications software, but also the systems software needed to make the computer systems run (12). There is, increasingly, evidence that the ability of climate-science applications to utilize distributed memory computers is limited (13). The human resources needed for the software effort are difficult to define, but it is on the order of the resources spent in the scientific effort. Given that, at the least, the development of high-performance applications software is extraordinarily difficult, the U.S. is in the position of needing to spend comparable dollars to those currently spent on scientific development in a high-risk activity. If successful, then it will take 3-5 years to develop comparable capabilities as presently available in other countries (14). However, long-term competitiveness requires availability of computational platforms with comparable usability and performance as those computers available to other scientists. To maintain a large software activity that our competitors do not have to maintain is a major fiscal inefficiency.

5.1) Software

Improved software and improved management of software lies at the foundation of the development of a high-end climate-science capability. The central role of software was introduced in Section 4.3. For the sake of organization the issues of software development will be subdivided into two major groups with the recognition that the two groups must be managed as a single entity with the primary goal to deliver a specific suite of products. This software infrastructure is represented in Figure 1 which shows the need to both isolate and layer activities so that they can be addressed in appropriate ways as well as integrating the activities towards shared goals.

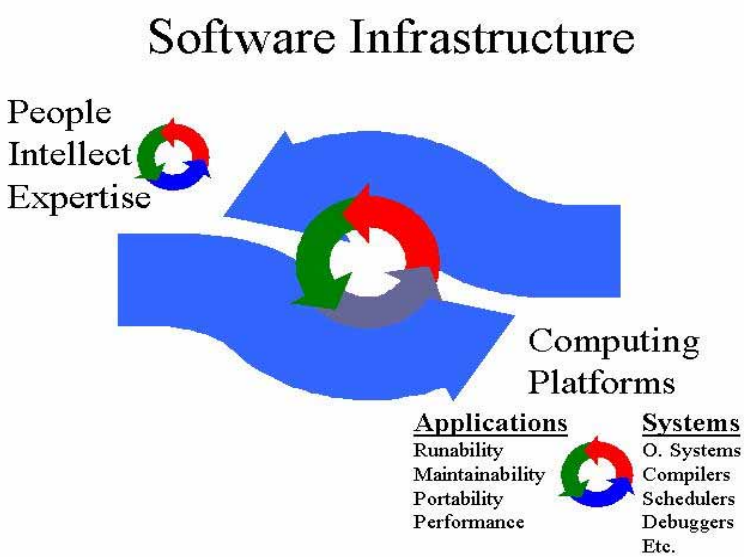

Figure 1: Schematic of software infrastructure. The purpose of the software infrastructure is to provide better integration and transition among the scientific and computational elements of the climate and weather modeling community. There are two major subsets of the software infrastructure. The first allows scientists at multiple institutions to work together concurrently in a controlled environment. The second allows computational and modeling issues to be brought together in an effective way.

5.1.1) <u>Software Infrastructure to Allow more Effective Scientific Collaboration</u>

Currently climate scientists at major U.S. centers work in both large and small groups that are organized around specific programmatic goals. The management of software varies in the different centers as well as within different groups within the centers. In most cases, individual scientists and their co-workers develop their software with significant independence and little attention to formalized software engineering conventions. In some institutions a group of software engineers exists to unite the development of the software for specific applications, in others, the software

is cobbled together on an as needed basis. Thus, major centers in the U.S. find themselves encumbered with large suites of difficult-to-manage software.

Nationally, the process that has been used to develop the software for major climate-science applications appears *ad hoc*. The impact of this on the ability of organizations to work together is large and negative. For example, consider the interaction of an individual university collaborator with a core activity at a national modeling center. The collaborator usually takes a version of the model and performs changes and experiments in a local research environment (15). After some time, the collaborator might have developed an algorithm suitable for incorporation into the core model. Then either the collaborator or center personnel have to take the candidate algorithm and redo the work of installing the algorithm and the testing with the new "current" version of the core model at the center. From the perspective of the university collaborator, there is a different interface with each modeling center (16). Therefore, to test the impact of, for instance, an algorithm for convective rainfall on weather forecasting, seasonal prediction, and climate data assimilation, the effort to make the collaborations can far exceed the resources spent on the actual scientific research.

The shortage of human resources leaves each modeling center with serious deficiencies in the development of an end-to-end climate-science capability. Therefore, more effective collaborations are critical to focusing adequate intellectual resources on a specific product-oriented problem. While the interaction between an individual researcher and a center, described above, might be fraught with manageable inconvenience, the interaction between centers presents insurmountable obstacles. Subtle algorithmic nuances arise that are, for instance, linked to an institution's history and increase the cost of the collaboration to overwhelm the perceived benefit of the collaboration. The ground is littered with well-intentioned collaborative efforts between individuals and between centers that have fallen victim to the overhead cost of collaboration. Virtually always, choosing to develop a given capability anew, internally, appears more cost effective than any benefit gained through collaboration.

A software infrastructure is needed that allows:

Concurrent development by multiple scientists at multiple institutions in a controlled environment (17).

Identification of a clear path of migration from discovery-driven research activities to core product-driven activities.

Partnering of software, hardware, and climate-science activities to assure optimization among computational resources, scientific quality, and scientific completeness.

It would be naïve to expect that simply defining a set of standards and guidelines for the construction of models might have a large impact on the integration of diverse modeling capabilities. The development of a software infrastructure requires the commitment to develop software management processes. These processes will require adaptation of the principles of software engineering to scientific development, with the focus on the end-to-end software system to provide specific prod-

ucts. The success of the software infrastructure will require commitment both from managers and practitioners. It is critical that software development processes be scaled to activities that include multiple principal investigators at multiple institutions. If this is not done, then there is little hope of developing the needed high-end climate-science capability.

5.1.2) Software to Allow more Effective Use of Computational Platforms

Traditionally climate scientists have worked with software specialists on code optimization. The realities of today's high-performance computing environment is that partnerships between software specialists and climate scientists are needed to assure that codes are capable of being run, maintained, and ported. That is, fundamental issues of software design and implementation have to be built in from the beginning, and computational decisions have to be considered on par with scientific algorithm decisions. If software issues and scientific issues are not treated concurrently, then the viability of codes for high-end modeling is in question. There are two major categories of software that need to be developed in order to provide a high-end climate modeling capability: applications software and systems software.

5.1.2.1) Applications Software

Applications software is the code that represents the scientific and statistical algorithms that are run to produce model simulations and assure their accuracy and merit. When vector computers were the work horse of scientific computing, scientists wrote code for a relatively stable computational environment that had well defined rules to enhance code performance. Software specialists from both scientific organizations and computer companies provided analysis of code performance and optimization. In many instances, possible improvements to code performance were isolated and tested by the software specialists and then provided to the scientists, who incorporated the suggestions if they deemed them worthwhile.

The move to commodity-based distributed computing (see Section 5.2) has changed the development of applications software dramatically. First, many of the tenets of traditional vector programming, which are deeply embedded in existing codes, are no longer valid. The hardware technology changes require going from successful vector software to more complex software for the distributed-memory computers. High performance of this new software is not assured. Second, the architectures towards which the applications are targeted are no longer stable. Different types of processors are being connected together with a variety of communications strategies.

The challenges facing the development of successful applications software are enormous. There has been significant investment in applications software at many U.S. institutions. Some applications have been successfully implemented. However, as the details of data use, schedule, and validation have been encountered in product-driven applications, a consistent high level of performance is not realized. There is now substantial evidence that the interprocessor communications requirements of many climate-science applications limit their actual performance on commodity-based machines to be much less than the theoretical performance specifications (18). In addition,

when considering end-to-end suites of application software, there are sequential processes and load imbalances that are in conflict with the parallelism of the problem.

We stand at the point of needing to develop techniques for software specialists and scientists to work together. Therefore, the infrastructure discussed in Section 5.1.1 must allow concurrent development not only by scientists, but also by software specialists. Codes need to be designed that partition the computational and scientific aspects of the code with well-defined, controlled interfaces. This should provide a more robust interface to the technology and allow adaptation to new technologies while buffering the impact on the scientific algorithms.

5.1.2.2) Systems Software

Systems software refers to that software needed to allow the hardware platform to be used. For this document, systems software will refer to a range of functions such as compilers, operating systems, debuggers, schedulers, and math and statistical libraries. These functions were supplied by vendors or third parties and were a purchasable commodity when Cray vector computers dominated the market. Today, the high-performance computing market is not large enough to provide vendor incentives to develop a robust suite of systems software on the premise that it will attract a specific customer base. Therefore, it is increasingly incumbent upon the applications communities to develop the systems software necessary to make their applications run.

Further complications come from the instability of technology. Increasingly there are interactions of application software with both the hardware and the systems software. Development of strategies to reduce these interactions and to accommodate technological changes is another challenge that must be faced.

5.2) Hardware/Impact of Technology Decisions

As stated above, the high-end computing industry in the U.S. has been transformed by the revolution in information technology over the last decade. Unlike other applications of information technology, the commercial potential of high-end computer, or supercomputer, systems is relatively limited. The vast majority of high-end computing systems are installed in government laboratories or academic institutions, with a smaller number being purchased by private industry for their research and development activities. With government support under the High Performance Computing and Communication (HPCC) Program, computer manufacturers in the U.S. were encouraged to design new architectures for high-end computers from mass-produced microprocessors. Microprocessor speeds were doubling every 18 months (Moore's Law), and it seemed reasonable that within a few years, that aggregate power could be applied to the most difficult computational problems. The new systems were envisioned to ultimately contain many hundreds to thousands of nodes, each containing a single processor or a cluster of a few processors, connected by a high-speed, integrated communications network. This new paradigm for high-end computing was labeled "massively parallel." The push towards massively parallel, distributed memory computing is still central to the policies of the Federal Information Technology for the Twenty First Century (IT^2) and Accelerated Strategic Computing Initiative (ASCI) programs.

While the U.S. aggressively pursued this strategy for high-end computing systems, the Japanese manufacturers NEC and Fujitsu continued the incremental development of parallel vector architectures composed of tens to hundreds of more powerful special purpose processors that share large regions of high-speed memory. The components of these high-end computing systems are designed specifically for scientific and engineering applications. Their development has been more evolutionary than revolutionary as the basic concept for these machines was pioneered by Cray in the early 1980s when it introduced its X/MP series supercomputers.

A Department of Commerce trade tariff ruling and political considerations essentially preclude the importation of Japanese high-end computing systems into the U.S. The legal test case for this policy was the acquisition of a computer system specifically for climate simulations by NCAR. While it is counterproductive to debate the costs and benefits of a policy that is unlikely to change, it is instructive to examine the differing perspectives of it that are held by the climate science and information technology communities. Not surprisingly, such an examination reveals a large divergence in objectives that has resulted in a culture gap between the fields. More importantly, unless the communities (and their sponsors) can agree on common goals, the prospect that investments in Information Technology research will benefit climate science is not good, despite the best intentions of well-meaning individuals in both camps.

5.2.1) Different Objectives Result in Different Metrics for Success

The pace of Information Technology (IT) innovation and application has been rapid over the last decade. Because the technology timescales are so short, the emphasis within the IT research community has been to take an idea quickly from concept through proof-of-principle and possibly the prototype phase. Coupled with this approach has been the tendency to extrapolate the near term technological trends into the future to drive the research agenda. Consequently, the metrics used to measure success within the IT community are based on the objective of *demonstrating the potential* of new technology. For high-end computing, the relevant performance metrics are based on three factors, theoretical peak speed, efficiency, and scalability. Theoretical peak speed is a hardware metric determined completely by the rated speed of the individual processors and the number of them that can be made to work together in a high-end system. Efficiency measures both hardware and software performance and is the fraction of that peak capability that can be tapped by an application on a given number of processors. Scalability is mostly a software metric that measures the increase in throughput rate that is achieved by an application as more processors are added to work on a problem. When combined, these three measures do provide an estimate of throughput performance. The tendency, however, has been to consider each of these factors as independent measures of progress. This philosophy is apparent in the well-known Top 500 (19) rankings that are published twice each year. These technology demonstrations have often been limited to a specific suite of applications software that only partially span the range of applications that require high-end capability (20). There is an implicit pre-selection of applications that are likely to perform well on these metrics, and the presumption that technology which has been proven on these applications will be generally useful. Actual performance numbers presented in Figure 2 show the wide discrepancy found in real applications.

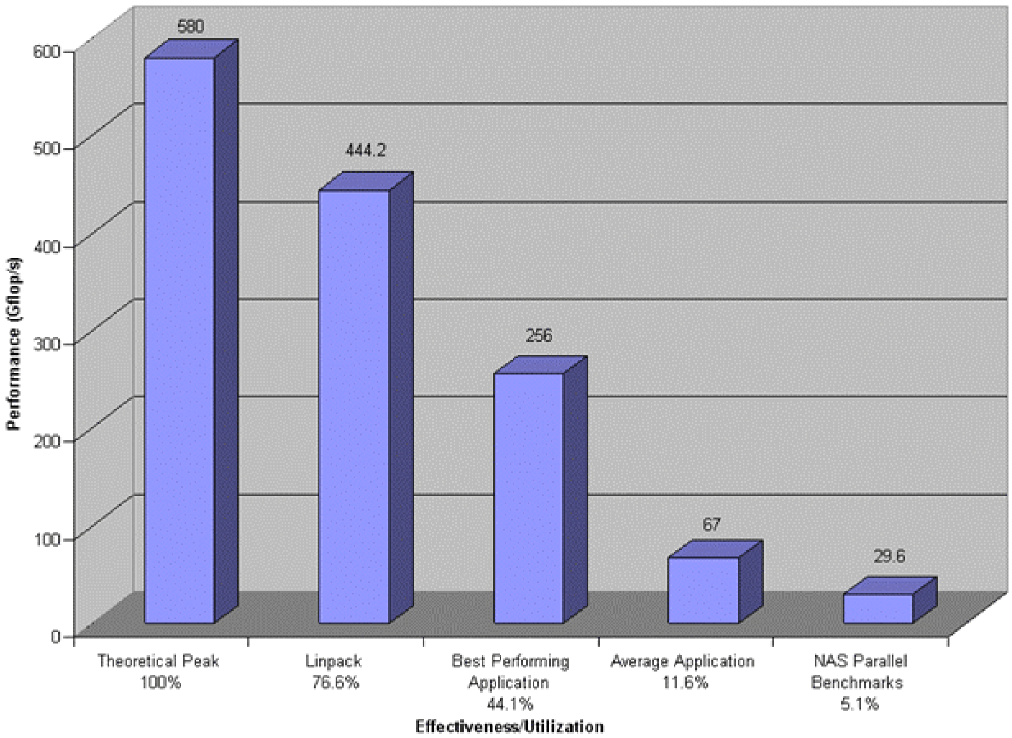

Figure 2: Measured performance of applications on the 644 processor Cray T3E at the National Energy Research Scientific-Computing Center (NERSC). This figure shows the sensitivity of performance to the specific application. The Linpack application is used to determine the performance reported in the TOP500 list. The TOP500 list is frequently cited to justify claims of computer speed, with the implication that this speed is relevant to a complete range of applications. These performance curves show that the Linpack benchmark does not represent a general application environment accurately. The average application at NERSC performs at 11.6 % of the theoretical peak. Within climate-modeling centers in the U.S. a 10% performance goal is often set. Numbers in excess of 33% are common on the Japanese vector computers, which coupled with their much higher single processor speed, leads to a performance-usability gap.

Over the same decade, climate researchers have seen their computational needs increase exponentially. While much of this demand has been met by the proliferation of powerful workstation technology, the most complete and sophisticated modeling experiments still require computing resources at the very high-end. Accordingly, the relevant computer technology metrics for climate scientists are measures of increases in capability, *i.e.* to run previously impossible or impractical simulations, and increases in throughput, which is a direct measure of model productivity. Many climate scientists assert that theoretical peak performance, efficiency and scalability are often used out of context, without explicit recognition that throughput is the critical measure for most applications. Based on throughput metrics, the climate science community has identified both a *performance gap* and a *usability gap* between the high-end computing technology developed in the U.S. with that developed in Japan. This performance-usability gap is represented explicitly in Figure 3, which is provided by ECMWF.

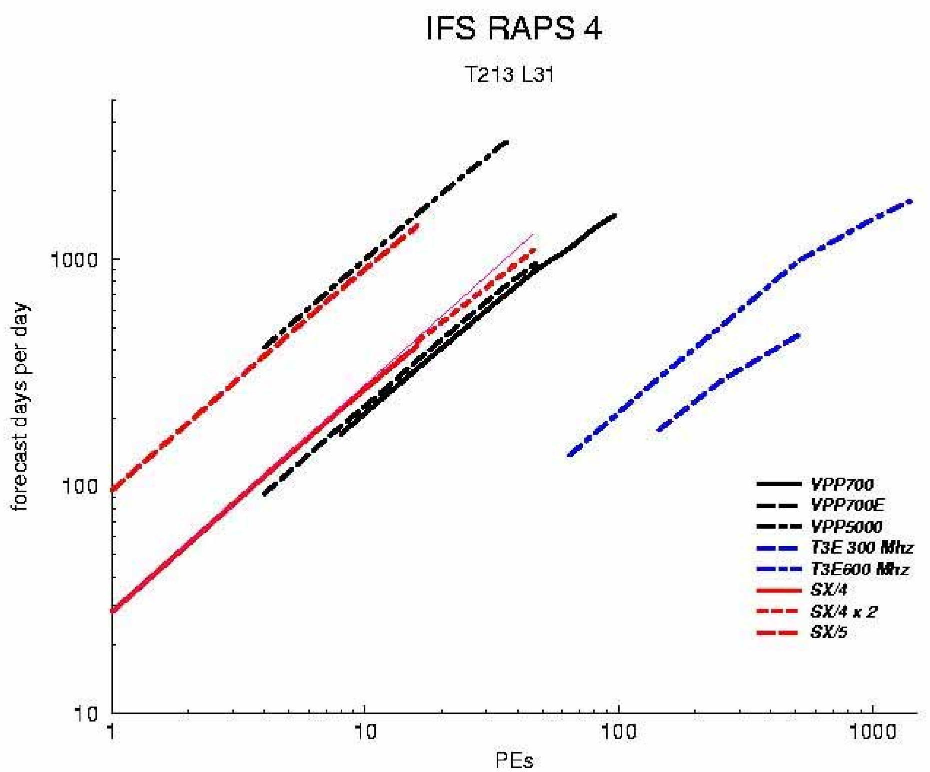

Figure 3: This figure shows the throughput of the forecast model component of the Integrated Forecast System (IFS). RAPS stands for Real Applications of Parallel Systems. This figure highlights the performance-usability advantages of the Japanese vector computers relative to the distributed memory T3E, which uses non-vector alpha processors. The VPP is from Fujitsu, the SX from NEC, and the T3E from Cray. The number of forecast days per computational day is plotted against the number of processor elements. Resolution is triangular truncation 213 (resolved waves) and 31 vertical levels. Comparable performance to the NEC and Fujitsu can be achieved *for the modeling component of the system* on the T3E. However, the T3E requires many more processors. Significant additional capability can be obtained on the NEC and Fujitsu by adding, for example, 10 more processors. From the graph, similar capability on the T3E would require more than 600 additional processors. Even if it is possible to obtain this performance, the software investment is significantly higher in the case for the T3E. Similar benchmarks have been run for the IBM SP systems and have yet to reach performance levels comparable to any of these systems. This indicates the difficulty of porting codes from one distributed memory machine to another. Figure provided by European Center for Medium-range Weather Forecasts.

5.2.2) Conflict between Climate Science and Information Technology

The NCAR computer system procurement was the seminal event that exposed the rift between two communities that previously had been cooperative and symbiotic. The NCAR benchmarks and acceptance criteria were based on the capability and throughput metrics described above (see, *Capacity of U.S. Climate Modeling to Support Climate Change Assessment Activities*, National Academy Press, 1998). The decision to buy a NEC computer system was met with shock and disbelief by many within the U.S. IT research community.

Although there have been several workshops and meetings, the groups appear to have been talking past each other since that time. On one hand, the IT community charges that the climate modelers have not embraced the potential of the new technology. Their argument continues that the climate modeling benchmarks are based on antiquated vector-friendly algorithms and codes. They further maintain that progress depends on the climate community making the investment in new algorithms and codes that are scalable and efficient on the new architectures, which the IT specialists foresee increasing by more than three orders of magnitude in the next ten years.

On the other hand, climate scientists maintain that despite ten years of development, the massively parallel designs have yet to become competitive with vector machines in many applications that require the movement of large amounts of information across processors. For example, the well-known NAS Parallel Benchmark suite achieves only 5% of peak theoretical performance on the Cray T3E at the National Energy Research Supercomputer Center, even though it gets nearly 77% of peak with the LINPACK Top 500 benchmark (see Figure 2). The climate community does, indeed, utilize modern codes, and there has been significant monetary and intellectual investment in the development of codes for parallel computers. While some of the developments have been successful, overall the performance has been poor on U.S. high-end computing machines. This is discouraging and does not motivate continued investment in seemingly futile undertakings. This view is reinforced when they see their international colleagues making greater progress while spending less time and money on software development.

While the Japanese high-end computing manufacturers have been stable and have followed a straightforward, predictable design path, there has been little stability in the U.S. high-end computing market. In the early 1990s, Intel Supercomputer and Thinking Machines dominated the market; neither is still in business. In the middle of the decade SGI, Cray (SGI subsequently acquired Cray) and IBM made product offerings that serve as the core of the U.S. installed high-end base today. Nevertheless, the workhorse Cray T3E line, which occupies half of the top twenty places in the current Top 500 list, has been discontinued and SGI recently sold its Cray division. Further, both Compaq and Sun unveiled plans in the last year to enter the high-end market. This turnover has led to usability problems, as immature operating systems, compilers, libraries and other software infrastructure components have undergone too few product cycles to be robust in a production-computing environment. IT researchers note that this competition and perceived turmoil is indicative of a healthy IT market and drives progress and innovation in the long term, which is true. It is also true, however, that the uncertainty and instability of the marketplace makes it difficult to build the robust production environment needed for climate research over the next several years when the science

and policy communities will demand both more and higher quality information from the modelers. The bottom line is: development of high-end computing platforms suitable for supporting product-oriented applications has become a research activity, with all the associated uncertainties and risk.

At the basis of the strategic decision to go towards massively parallel computing was the assumption that the incremental development of vector computing platforms had reached the end of the line. This decision was as much economical as technological. The evolutionary development of vector machines in Japan followed a strategy that U.S. vendors (*i.e.* Cray) had rejected. There are indeed economic considerations about the viability of the Japanese vector computers. Fujitsu will no longer market vector machines, leaving NEC as the lone vector supercomputer. No matter what the ultimate fate of the vector computing business, however, the Japanese vendors have already delivered ample computing platforms to assure the gap between U.S. and non-U.S. climate-science centers will increase for at least the next five years. The emergence of a new U.S. effort centered around the Tera purchase of Cray from SGI and the formation of Cray Inc. does little to improve the situation in the short-term. The long-term effects are beyond our ability to evaluate. There is some optimism as large-cache fast processors are finding a broader market, which will allow substantial throughput to be achieved without requiring scaling to many hundreds of processors.

5.2.3) Solution is to Re-establish Cooperation Around Common Objectives

If the U.S. is to remain among the intellectual leaders in climate modeling, there needs to be recognition that the issues of both the climate community and the IT community have merit. The IT vision is clearly long-term, and it is not clear that high-end climate community can make it through the near-term crisis and maintain an intellectual critical mass. There has to be directed investment to take the proof-of-concept activities of the IT community and develop stable hardware and software environments suitable for supporting product-oriented activities. The technology experts and the science experts need to agree on a common set of both near-term and long-term objectives, then develop a workable strategy to achieve them. One possible solution is to look back to 1960s and 1970s when scientific supercomputing first established itself as a necessary tool for science and engineering. Discipline-oriented centers such as NCAR and GFDL became magnets for mathematicians and computer engineers who were attracted by the challenges of improving the primitive technology and solving real world problems. As a result, these centers became hotbeds for supercomputing research that had utility beyond the immediate applications. These centers would complement the multi-discipline, many-user supercomputing centers, as they could tailor their configuration and management to the needs of a smaller and less diverse research community. Proper sponsorship of these centers requires that the current friction between the communities be replaced by a healthy and cooperative alignment of both scientific and technological goals. Strong, innovative and forward-thinking leadership will be required for this approach to succeed.

5.3) Characteristics of climate-science computing

Given the inconsistent success of executing climate-science algorithms on distributed memory parallel computers, it is worth noting the parameters that define the computational problem. As with many scientific applications, climate-science problems require fast computers as well as high capacity, fast-access mass storage systems. Compared with other scientific applications, climate-science data impact the computational problem in both direct and indirect ways. Indirectly, the long history of weather observations has led to the development of complex physical parameterizations that represent subscale processes. These parameterizations are not executed uniformly across the discrete domain of the models; that is, their execution is dependent on local conditions. The results of their execution might then need to be communicated to other parts of the geophysical (and computational) domain. In total, the demand to represent localized physical processes in global models introduces difficult load balance and communications problems that reduce their potential to scale to many processors.

Data usage, also, has a number of direct impacts. Climate-science data require high capacity storage and are often heterogeneous in their format. Assimilation of these data brings a whole new level of difficult challenges. Assimilation algorithms are, themselves, often more computationally intensive than climate models. The assimilation process interrupts simulations for the data insertion, which requires overhead as the process starts and restarts after every few hours of simulation. In its routine use of data assimilation, weather and climate science is unique. Currently, all relevant organizations are having difficulty achieving massive scalability of assimilation systems.

There are other external factors that have a profound impact on the computational environment. First, climate-science products often have to be delivered on schedule in order for their utility to be realized. While this is obvious for weather forecasts, time-critical requirements arise in climate and chemical assessment activities. Time criticality directly impacts the capability requirements of the computational platform and focus performance metrics more definitively on throughput rather than processor speed. Second, the impact of weather and climate is on regional and human scales. Therefore the results of global modeling activities are required to have precise information on scales that are much smaller than can be directly simulated.

Computational activities to address climate-science computing need to consider all of these aspects of the computational problem. This requires significant attention to software engineering, systems engineering and systems design, which is outside of the scope of the research programs that normally support Earth and computer science.

6) <u>Human resources</u>

Managers in all the Agencies are confronted with a number of human resource problems (21). There are several funded activities at both government laboratories and universities who are seeking climate scientists to fill similar positions, for example, in the development of the next generation of physical parameterizations. Qualified scientists to fill these positions are rare, with advertisements often drawing no candidates with sufficient expertise. Furthermore, in recent months, both Japanese and European Centers are recruiting prominent U.S. scientists. U.S. science organizations

are often dependent, already, on immigration of foreign students, and the presence of strong foreign centers attracts these students away from the U.S. It is clear that there are not adequate scientific personnel for all of the currently existing groups to develop comprehensive capabilities. Further, currently existing activities must be the basis of any timely, strategic solution to address the gap between needs and existing capabilities.

There is another significant pressure on the human resource pull that comes from the booming information technology economic sector. Information Technology professionals who are needed to complement scientific personnel are attracted to higher paying jobs in non-scientific businesses. Science-related computer jobs are increasingly viewed as a niche profession and do not attract career-oriented computational experts. Most U.S. climate-science centers are seeing increased turnover in computational positions, with a net migration away from the field. Significant numbers of Earth scientists are leaving the field after school, before accepting a first scientific position. In order to compete with the non-scientific information technology job market, scientific organizations need to offer not simply competitive salaries, but development of job skills that are attractive to mainstream professionals and career paths comparable to the field as a whole.

7) Management/Business Practices/Institutional models

7.1) An "Institute" for Product-driven Climate Science

In order to provide the required capabilities and products there needs to be an organization that has as its primary mission the delivery of these products. The current approach of expecting the existing organizations to deliver these products as an ancillary activity to their primary missions is not working. Simply sprinkling these organizations with additional funding to give them incrementally greater capability is not an effective remedy to the current situation. The attributes of the institute to deliver these needed products will be discussed next. This is done without consideration of which Agency should host this institute or, indeed, if a completely new institute external to all of the agencies should be initiated. We, ultimately, envision the evolution of a Climate Service that integrates all aspects of Modeling, Data, and Computational Systems. However, the planning and development of this service is difficult and necessary and subject to reconciliation with current Agency missions. Therefore, we propose an evolutionary process that starts, soon, to align the major components of a climate service while a more structured service is planned and developed.

Success of a product-driven climate service requires fundamental changes in Agency behavior and the discovery-driven science culture. Fortune 100 companies today have realized that it is fruitless to try to graft cutting-edge business theories and new agendas onto the framework of old and unproductive organizational practices. Moreover, the success of organizations is usually traced to the strength and commitment of an individual or a handful of individuals in key positions. Therefore, those in top positions must exhibit strong and stable leadership as they incorporate the positive qualities from the existing programs while not losing sight of the need to initiate and implement the overarching vision for the new institution.

7.2) Institutional Attributes

First and foremost the institute charged with the delivery of needed climate-science products must have a clearly defined mission focused on the actual delivery of the product, including fundamental assurance of the quality of the products. Solid scientific process must lie at the basis of the products, but it must be realized that delivery of the product will require bringing closure to incomplete scientific arguments to allow builds of software suites. These builds need to be tested and validated prior to their application in product generation and will serve as baselines for future builds with more comprehensive scientific development.

The defined mission of the institute will provide overarching structure to facilitate prioritization and decision making. Just as essential as the mission, there needs to be an executive decision-making function vested in, at most, a small group of science and software managers, whose performance is measured by the successful delivery of the products and the subsequent customer response. At the lead of this group will be an individual with the ultimate authority and responsibility for delivery of the product.

For the executive function to be effective, the institute has to have a unifying incentive structure that connects the organization from top to bottom, with the delivery of a successful product at the top. The current situation does not support an effective incentive structure at any level. At the lowest level, scientists are generally rewarded for individual accomplishments of innovative research: *i.e.*, discovery-driven research. At the next level, even in the most project-focused organizations, funds flow into organizations from a variety of program managers. These funds fuel subsets of the organizations and draw human resources towards these subsets, away from the systematic delivery of the needed products. The program managers naturally command the allegiance of these subsets of the organization and are generally not rewarded for the delivery of successful products by the organizations they fund. This programmatic fracturing extends to computational resources, and in most U.S. laboratories there is a disconnect between computational resources and the delivery of simulation and assimilation products. Finally, the organizations that are expected to deliver the needed climate-science products are often embedded in large Agency laboratories whose basic metrics of success do not include delivery of successful climate-science simulation and assimilation products. All told, the current structure of climate-science activities in the U.S. is fracturing rather than unifying.

7.3) Business Practices

A functioning Climate Service that contains the attributes described above would stand in stark contrast to the pervasive scientific culture of the U.S. Such an organization would vest the decision making function in an executive process that acts in the best interest of the delivery of the institutional products (23). Such a Climate Service will require supporting business practices that are significantly different from those currently used in the scientific community. These business practices must be unifying. They must provide a mechanism for stable and effective external review as well as integration with the discovery-driven research community.

As with the scientific and computational aspects of this enterprise, the business practices need to be considered in a systematic and integrated way. They need to support the goals and function of the charged institute. While the complete specification of these business practices are beyond the scope of this document the following can be derived from experience within the current organizations.

Funding must be
- focused on delivery of products
- stable
- balanced on all elements of the organization
- under the direction of the executive-decision-making function
- isolated from program volatility of funding agencies

Review
- conventional peer review will not work
- need to develop review techniques to support organization
- different levels of review are needed for scientific and programmatic purposes

Business practices
- success of the climate service must be a critical metric for success of the hosting agencies
- contractual vehicles must support the organizational goals
- salary structures to allow effective recruiting and retention of personnel
- etc.

8) Recommendations

Recommendation 1) Formation of a Climate Service: A Climate Service with a well-defined mission should be chartered to deliver simulation and related data products. This should be a national resource that allows investigation in a controlled scientific environment with well-defined interfaces to both research communities and non-scientific customers.

Recommendation 2) Build from existing expertise: Given the human resource limits and need to address climate problems in a timely way, the Climate Service must be built upon existing expertise, with clear separation of Climate Service functions from current Agency obligations. The quality of discovery-driven climate-science research in the U.S. is excellent and broad, and a new Climate Service should strive to strengthen and focus the research activities on important national priorities.

Recommendation 3) Location: The Climate Service should not be located or assigned to any Agency or Center within the current multi-agency framework. We propose that an independent service, which is a concerted federation of the appropriate current Agency capabilities, should be formed. The existing agencies need to act like member states, drawing from a concept successfully used in the European Union (24).

Recommendation 4) Business and Management: An integrating management structure with an executive decision making process and a supporting incentive structure must be developed. Supporting business practices must be developed. The external review and oversight process needs to be developed to allow stability and insulation from short-term programmatic volatility. Without a new business model, incremental funding of existing organizations will not provide needed capabilities.

Recommendation 5) Computational systems: The Climate Service requires dedicated computational resources with the highest level of capability. The computational resources must be aligned with the generation of the Climate Service products and under the management of the Climate Service. If additional computer science or computational science research is needed, then this, too, needs to be managed consistently with the goals of the Climate Service. This research must be application driven with the development of integrated software and hardware platforms to deliver the Climate Service products.

Recommendation 6) Software: Of the major Elements of a climate-science organization, Modeling, Data, and Computational Systems, the financial investment in the software component of the Computational System Element is most deficient. Many of the software issues are straightforward and require the integration of software engineering personnel and principles into climate-science activities. A software infrastructure needs to be developed to allow multiple groups at multiple institutions to participate in concurrent development in a controlled environment. Similarly software investment is needed to allow better interface with the computing environment – facilitating runability, maintainability, portability and performance. The community must also develop a strategy that

supports development of systems software in the absence of provision by vendors.

RECOMMENDATION 7) Hardware: Near term investment in hardware should be targeted to provide incremental increase in capacity, allowing the exploitation of task parallelism – running multiple copies of application software.

Long-term investment in hardware is dependent upon many uncertainties in the development of applications software and technology development. Significant systems engineering and design is required to reduce the risk in technology investment. It is incumbent to maintain maximum flexibility, including the development of software that buffers the volatility that arises from changes in technology development.

> The U.S. policy on high performance computing adversely affects the climate-science community and places U.S. centers at a competitive disadvantage with centers in other countries. At the least, the usability and performance of Japanese vector computers already delivered to non-U.S. centers assure that by many metrics the U.S. will lag non-U.S. centers for five years. There is substantial risk that this deficit will last longer with pervasive negative impact on U.S. climate-science.
>
> The importation or offshore use of Japanese vector computers would have a profound near-term impact on U.S. capabilities – decreasing the gap between U.S. and non-U.S. centers.

RECOMMENDATION 8) Software-hardware dependencies: The long-term investment in hardware and software should be approached in a systematic way that focuses on the software suites of the Climate Service. A hardware-centric approach that focuses on the placement of computers simply defined by high theoretical processor performance will not be effective.

> The software investment depends greatly on the hardware options. With the restriction of U.S. centers to distributed-memory, commodity-based processors a large investment is needed in applications software, and there is substantial risk that the performance of the software is intrinsically limited. This investment is significantly higher than that necessary for non-U.S. centers. If Japanese computers were available to U.S. centers, it would not serve as substitute for significant investment in software infrastructure.

RECOMMENDATION 9) Integration: There should be the formation of a Climate Service that is roughly parallel to the National Weather Service. Earlier reports (see reference list) suggested the formation of, perhaps, 1-3 climate-focused activities as well as fully evolved capabilities in seasonal to interannual prediction. Issues of chemistry and data assimilation were not addressed in these earlier reports.

We recommend two major core simulation activities. The first is focused on weather and should build from the National Weather Service. The second is focused on climate, and while it builds from

existing expertise, the exact components and location of the Climate Service is dependent upon addressing a number of organizational and management issues discussed throughout this document. With strong leadership and a clear vision, many of the principal components of Climate Service can be aligned, in the near term, from existing product-oriented activities in the mission agencies. It is critical that initial steps be made to develop a credible and competitive high-end climate capability, and we are concerned that potential Agency and political positioning over the location and running of a potential Climate Service will delay its formation.

The Weather Service and the Climate Service should undertake the development of a unifying infrastructure to allow effective transfer of expertise and algorithms. There is significant potential benefit from more thorough integration of weather and climate activities, as well as numerous other sub-disciplines that have developed to significant maturation as a field. On the time scale of ten years, a useful vision is one of a unified national capability for weather and climate modeling with focused centers for specific applications.

With this call for a more managed product-driven Climate Service, we maintain that a robust and diverse discovery-driven research capability must be sustained. If "operations" were to come to dominate the entire climate-science community there would substantial negative impacts. A balance is needed between product-driven and discovery-driven activities, which each benefiting the other. Due to the complexity of both the scientific and programmatic aspects of these fields, integration and mergers will be difficult to manage.

> The issues of integration are amongst the most difficult to reconcile. There are many difficult conflicts that must be considered when deciding what to include in a climate service, and/or when to include it. Of the underlying issues there are two that directly impact our recommendation: 1) the shortage of human resources, and 2) the ability to support an integrated product-driven capability in the current multi-agency culture.
>
> The shortage of human resources motivates tight integration as there are not enough people to populate several comprehensive modeling efforts. The leadership, management, and business challenges motivate keeping the implementation as simple as possible and expansive integration increases the risk of failure. The need for simplicity suggests, therefore, that the nascent Climate Service should focus only on climate simulation and that chemistry, data assimilation, carbon cycle modeling, etc., should not be included. However, these other activities would continue to be advocated and developed to meet both programmatic and scientific needs. This would place them in direct competition for human and monetary resources as well as further entrenching the fragmenting processes that are already in place. Further, efforts in other countries are already including chemistry and data assimilation in their climate activities and if a nascent U.S. service did not, then there would be a competitive disadvantage from the beginning. This conflict between simplicity and inclusion of activities broader than traditional climate simulation is complex and requires more deliberation than possible in this document. Further, some strategic positions

need to be taken at Agency and higher levels on what should be included. We make the following observations.

Seasonal-to-Interannual Prediction: The issues of how to fit seasonal-to-interannual activities together with longer term focused, *i.e.* decadal, centennial, *etc.*, are amongst the most controversial. Seasonal-to-interannual activities are on the brink of operational utility, and could be viewed in several ways, for instance, an entity unto itself, an extension of operational weather forecasting, or the foundation of a climate service. There is clearly a strong link between any climate service and seasonal-to-interannual activities, as one metric of a credible climate model will be its ability to perform seasonal-to-interannual predictions. If operational capabilities are truly on the brink, then alliance with NOAA operational activities is called for. NOAA has the primary operational capability in the nation and, whether or not this operational capability is as robust as necessary, the generation in the near-term of an independent operational capability is not justified. If the Nation is to generate additional operational capabilities, because the current capabilities are deemed inadequate, then we feel this needs to be done in an integrated systematic way that would benefit not only seasonal-to-interannual activities, but also other climate activities and, in fact, should also address those perceived inadequacies in the current weather operations (25). There would remain the need to provide seasonal-to-interannual expertise within the proposed Climate Service as well.

Chemistry, Data Assimilation, etc.: Chemistry and data assimilation are fast becoming important aspects of the climate problem. There is legal mandate for ozone assessments, and the Nation is struggling to provide the infrastructure to provide these assessments. Foreign centers are increasingly incorporating these more comprehensive capabilities in their climate programs. Ideally a climate service capability would include both a chemistry and data assimilation capability from the beginning. However, we feel that these capabilities need to be brought in incrementally. Chemistry capabilities should be brought in sooner rather than later. The programmatic and scientific issues of data assimilation are more complex and inclusion in a Climate Service at the time of formation increases the organizational complexity too much. We express concern that initiatives to develop capabilities in carbon cycle modeling and weather research compete for the same resources as the climate community.

Again, the issues of integration are complex and beyond the scope the current document. If the tenets outlined in this document are accepted, then we propose that the issues of integration be considered in the implementation planning. Success would require statement of strategic desires at the Agency level and above, as well as inclusion of the practitioners in the affected communities.

Recommendation 10) Budget: On the order of 150 scientists, software engineers, and application-directed computational scientists, programmers and computer scientists need to be dedicated to the Climate Service. This number is dependent on the level of integration that is sought in the Climate

Service (see Recommendation 9). This would require approximately $20M.

The money for computing systems is more difficult to define. We propose that on the order of $10-12 M per year needs to be allocated for hardware and related services for high-performance platforms. Similar amounts of should be allocated for other computational capabilities and services within the Climate Service (26). Thus the total cost for computational resources is $20-24M. This is a lower limit.

1.5 million for facilities

$7.5 million to integrate with research and customer community

Total, approximately $50 M

This capability cannot be built by redirection of research funds that are linked with the USGCRP and will require infusion of new funds. The "mission" agencies, NOAA, NASA, and DOE, already fund some activities that are product oriented. NSF, through the Community Climate System Model Program, has taken on the responsibility to provide a research facility for the climate community. These existing capabilities could contribute to the core of a putative climate service. This requires the Agencies to clearly distinguish the funds that might support this core from those funds that support their discovery-related research and manage them appropriately. With the consideration of these existing product-oriented activities, which are broader than those just associated with USGCRP, the needed capabilities can be built from a combination of existing and new funds. The new funds must support the development of product-driven activities and not simply appear as enhancements to existing research activities. Finally, the current discovery-driven research programs generally support high quality and important investigations; therefore, to build a product-driven research institution at the expense of the discovery-driven programs would undermine the underlying research environment that is broadly cited as the greatest strength of the U.S.

RECOMMENDATION 11) Implementation: The details of implementation will require significant planning and be dependent on a number of interrelated decisions that must be made by the Agencies. Strong leadership, both within the Agencies and at a level higher than the Agencies, will be required.

The implementation can and should be incremental. In fact, we believe that with the definition of a stable vision and leadership there are a number of existing activities that could form the core of a future Climate Service. There are already moves by all of the Agencies to better integrate and unify modeling and computational activities. If these can be orchestrated towards a long-term vision, then substantial steps can be taken while the details of the Climate Service are developed and evolved. Again, without a new business model and management strategies within which to organize the Climate Service, there is a danger of simply rearranging the current activities, which will not be successful.

Finally, we emphasize that is artificial to speak of a climate-science capability, a national climate service, without integration of modeling and data (*i.e.* observational) activities. As charged, we addressed the data activities, but they were not explored in as much depth as the modeling activities. We state, explicitly, that many of the same underlying problems affect the environmental data undertakings of the U.S. as affect the modeling community, and integrated, systematic solutions are ultimately needed. Additional funding is crucial to both develop foundation climate observing systems and to integrate and maintain existing data sets for climate applications.

9) **Reference documents**

National Academy Reports

Capacity of U.S. Climate Modeling to Support Climate Change Assessment Activities (National Academy Press, 1998)

Adequacy of Climate Observing Systems (National Academy Press, 1999)

Previous Reports and Reviews on U.S. Modeling

The Accelerated Climate Prediction Initiative: Bringing the Promise of Simulation to the Challenge of Climate Change (Pacific Northwest National Laboratory, June 1998)

JASON Review of Advanced Climate Prediction Initiative (S. E. Koonin, Reporting Author, August 10, 1998)

Report on the NSF/NCEP Workshop on "Global Weather and Climate Modeling" (D. L. Williamson, NCAR, Lead Author, 24 February 1999)

Recommended Implementation of Accelerated Climate Prediction Initiative (ACPI) (W. L. Gates, Lawrence Livermore National Laboratory, Lead Author, April 1999)

Reference on data usage

Errico, R. M., Workshop on Assimilation of Satellite Data, *Bull. Amer. Met. Soc.*, 80, 463-471, 1999

Suggested reference on organizational and business models

Senge, P. *et al. The Fifth Discipline Fieldbook: Strategies and Tools for Building a Learning Organization* (Doubleday, 1994).

References on computational issues

Christensen, C. M., *The Innovators Dilemma*, Harper Business, 2000

http://www.top500.org/
Compilation of the top 500 supercomputer facilities in the world.

http://www.hpcc.gov/ac/testimony/kennedy6oct98.html
Testimony of Professor Ken Kennedy to Congress

10) Endnotes

1 For example, *The Washington Post*, June 12, 2000, pA03; June 14, 2000, pA38.

2 The term curiosity-driven is also possible.

3 To be clear, there are also high-end modeling activities that support discovery-driven research. This is the traditional role of the Community Climate Model at NCAR. This particular example does, in fact, have a product, the provision of a documented comprehensive model to the research community. In this case a facility is being provided to the research community.

4 http://www.ndsc.ncep.noaa.gov/

5 *i.e.* massively parallel

6 It is important to distinguish between computational capability and capacity. Capability is the execution in a given wall-clock time of a job requiring the entire computational platform, and capacity is the aggregate of all the jobs running simultaneously on a platform. Increasing capacity allows the execution of more jobs of the same size. Increased capability is required to run larger jobs, *e.g.* more resolution, more comprehensive, or to achieve faster throughput of a same-size job. Increased capability requires software development in order to utilize the potential capability of the hardware. Increased capacity is easy to buy. Increased capability is hard to build.

7 Again, similar issues exist in the Data Element.

8 Further, a conscientious effort needs to be made to align resources and incentives with these efforts, as well as to separate the integration activities from short-term programmatic goals. It must be recognized that the incentives to many individuals to advocate change to a more integrated climate service is low. By in large, scientists have been successful and autonomous, and, by definition, movement to more product-oriented research removes some of their autonomy. Furthermore, it is easy to point to numerous government-sponsored activities to 'centralize' computational systems, data activities, or modeling efforts that have either fizzled or failed. Therefore, the trust in management is low. Also, since the ultimate goal of an effort to integrate modeling activities will consume resources, some see a direct threat to their research activities as the product-generating model takes precedent over individual activities.

9 *i.e.* massively parallel

10 see Innovators Dilemma, in the reference list.

11 see Ken Kennedy Testimony, in the reference list.

12 There is little evidence that the market for high-end computing is adequate to motivate vendor development of easy-to-use systems software

13 *i.e.* scalability is limited

14 A tremendous uncertainty in these arguments is the viability of the Japanese supercomputing industry. Already delivered Japanese-made processors assure that U.S. scientists will remain behind for the next 3-5 years. With the expected continued development and manufacturing of specialty Japanese processors, the gap will widen.   As long as U.S. computers are based on relatively slow processors in distributed memory architectures, and there are relatively fast Japanese processors with shared memory architectures, U.S. scientists will not catch up. Increased capability will be much easier to achieve on the Japanese-processor based computers.

15   The environment remains "local" even if the university collaborator uses a computational facility at the collaborating center.

16   Science and Program Managers need to be fully aware that the bulk of the computational resources in a climate-science center go for testing and validation.

17   Controlled in both the scientific sense of controlled experimentation, as well as controlled in the sense of there being standard processes in the development of configured software suites for specific applications.

18 see Ken Kennedy Testimony, in the reference list.

19 see Top 500 web site, in the reference list.

20 see *Evaluating System Effectiveness in High Performance Computing Systems* by AT Wong, L. Oliker, WTC Kramer, TL Kaltz and DH Bailey Lawrence Berkeley National Laboratory Draft Report 44542, November 11, 1999, available in PDF format from http://www.nersc.gov/aboutnersc/pubs/esp.pdf

21 for example, *The Washington Post*, May 8, 2000, A21.

22 This arises because the computing organization is often funded to pursue computational research within information technology programs or the computing facility is run as an institutional facility and the product generation exists in an uncomfortable balance with large numbers of small discovery-driven research projects.

23 This focus on individuality manifests itself in large projects through the development of consensus decision making. It becomes necessary for all members of the project to have ownership of the project. Therefore, such projects are generally driven by the development of individual capabilities with only collegial or casual integration towards project goals. The needed infrastructure, including

process definition, is generally underdetermined and under funded. This leads to the project taxing the, generally, already stressed schedules and resources of the individual participates, who are generally not directly rewarded for their participation in the project.

24 If the decision is made to locate the Climate Service in an existing Agency or center, then the metrics of success of the Agency or center must be realigned with the success of the Climate Service a critical metric of Agency or center success. Agency mission and business practices would need to be altered to support the delivery of the products of the Climate Service.

25 Operational capabilities rely on data access. There is much in common with climate and weather data, and shared data infrastructure would seem to be at the core of any and all operational capabilities.

26 ECMWF spends approximately $9M per year on the service for their Fujitsu computer. Using the numbers from the Department of Commerce tariffs would suggest that at least $45M per year would be needed to obtain similar capabilities with U.S. computers. Therefore, substantially more money might be called for. However, expenditure of such large amounts of money in currently U.S.-available hardware with currently available software is unjustified.